A PROGRESSION FOR THE PREDICTION OF EARTHQUAKE USING DATA MINING HAVING AN INFORMATIVE IMPACT ON SOCIETY

SHORT SYNOPSIS

Submitted in partial fulfillment of the requirement of the degree of

DOCTOR OF PHILOSOPHY

to

Manav Rachna University

by

Ms. Sumita Choudhary
2K17-PFE-014

Under the supervision of

Dr. Prinima
Associate Professor, Department of Computer Science &Technology
Faculty of Engineering

Prof. Felix Musau
Dean of School of Computing, Riara University, Nairobi, Kenya

MANAV RACHNA
Vidyanatariksha

MANAV RACHNA UNIVERSITY
(FORMERLY MANAV RACHNA COLLEGE OF ENGINEERING
NAAC ACCREDITED 'A' GRADE INSTITUTION)
Declared as State Private University under section 2f of the UGC act, 1956

Year 2021

MANAV RACHNA
vidyaparikshan

MANAV RACHNA UNIVERSITY
(FORMERLY MANAV RACHNA COLLEGE OF ENGINEERING
NAAC ACCREDITED 'A' GRADE INSTITUTION)

Declared as State Private University under section 2f of the UGC act, 1956

DECLARATION BY THE SCHOLAR

I hereby declare that the shortsynopsis"**A PROGRESSION FOR THE PREDICTION OF EARTHQUAKE USING DATA MINING HAVING AN INFORMATIVE IMPACT ON SOCIETY**" beingsubmitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy in department of Computer Science & Technology under Faculty of Engineering of Manav Rachna University, Faridabad during the academic year 2021.

It is a bona-fide record of original work carried out under the supervision of **DrPrinima** and **Prof. (Dr) Felix Musau** has not been submitted in full or part to anyother university or institute for degree or diploma. Further, it is certified that short synopsis is within the plagiarism limit and copy of the same is attached as per Universitynorms.

Ms. Sumita Choudhary
2K17-PFE-014

Signature oftheSupervisor
Full name, Designationand Faculty

Signature of the Co-Supervisor
Full name, Designation andFaculty

PROF. FELIX MUSAU
DEAN, FACULTY OF COMPUTING SCIENCES

1/10/21

Chairperson, DRC

Contents

S.No.	Description
1	Abstract
3	Introduction
4	Literature Review
5	Problem Statement
6	Determinable Objectives
7	Methodology
8	Conclusions
9	References

Abstract

Earthquake is a major debacle and entire world, and researchers are worried about this calamity. This calamity not only ascertains in injury, fatality of lives, and damages of buildings, property but also makes people helpless and homeless. This in turn, brings fright and forces people to concern about the causes of earthquake. The consequences of earthquake occurrences in near future are a dispute on lives of people globally. Earthquake forecasting is the demand of human race and a complex process to analyze dependency of various factors on the occurrences of the events. Precursors observed during earthquake can be treated as a parameter and through the studies and techniques of data mining; models can be prepared to show a correlation on the precursors and parameters. To analysis the problems and encourage the researcher these developed models are ideal to find essential solution to inherent in the prediction of earthquakes by applying the right data mining techniques. Magnetic fields and electric fields, change in level of ground water, change in temperature, radon gas emission, uncommon gesture and signal among animals of the specified area not only requires observation but must be analysed and explained. This analysis explores knowledge on the inter dependency of these possible forerunners along with the nature of occurrences of earthquake having specific location and time. Studies have shown that precursors have shown similar properties at the epicentre of the earthquake. In many incidences, these forerunners are clearly intimidated just before the occurrence of strong earthquake. This study and analysis is suggesting a deterministic and investigative approach for earthquake prediction thus strategized and planned models like time series, clustering, regression, decision tree, Bayesian model, Neural network etc. are considered for better result. Published scientific literature and catalog were contemplated and accepted to distinguish earthquake forerunners and determiners. There are small sized and measured earthquakes which happen in the same place first and then later or afterwards the large-scale earthquakes follow and occur at the same place thus raising and recognizing the major shock of the collision of the plates of the earth. An earthquake is a foreshock cannot be thumbed by researchers unless the bigger earthquake takes place. The main earthquake which can be larger than others is only recognized by the major shock and coined as primary earthquake. After shocks are the prominent attributes to follow the major shocks. These occurrences of earthquakes whether small or large in nature depend on by the energy released suddenly in a specified region of the rocks of the earth and recognize main shock of the sliding movements of plates. For longer time like weeks, months, and even years after the major shock aftershocks of the

earthquake can continue to happen. The algorithm based on statistical model can be mapped out by employing a simple linear equation, or it may be a complex neural network, or by modern software Natural Language Tool kit, GNU octave programming. The Time series model using Python language is to co related with data mining techniques like, regression analysis, time series in pattern evaluation. Based on pattern recognition a LSTM method is designed to derive a logical algorithm on different earthquake parameters and precursors to quantify the probability of an earthquake occurring within specified Time, Location, Depth, Magnitude, variations on Temperature, Pressure and Animal behavior.

Keywords: Tectonic plates, Seismic waves, Magnitude, Decision tree, Neural networks, LSTM, Random Forest

1. Introduction

There are four prime coverings of the earth - the inner layer, outer layer, mantle and crust. The exterior part of our planet earth has a thin skin covering the prime and main part of the mantle and the specific part of the crust. Though this skin encompasses pieces in adequate numbers like a confused state encloses the exterior part of the earth but it is not found as one piece but in all the parts experienced on the crust of the earth. Moreover, these confused pieces slide past one another, keep moving slowly around and collide with each other. These tectonic plates are nothing, but the baffled and confused pieces and the plates borders are represented by edges of the plates.

Almost all the earthquakes across the world occur on these plate borders which in turn, show many faults. As the edges are uneven in nature, thus, the remaining part of the plate keeps moving and the edges of the plate get stuck to earth's crust. Earthquake finally, occurs when the edges detach from one of the faults and plate has moved far enough with an impact from the earth's crust.

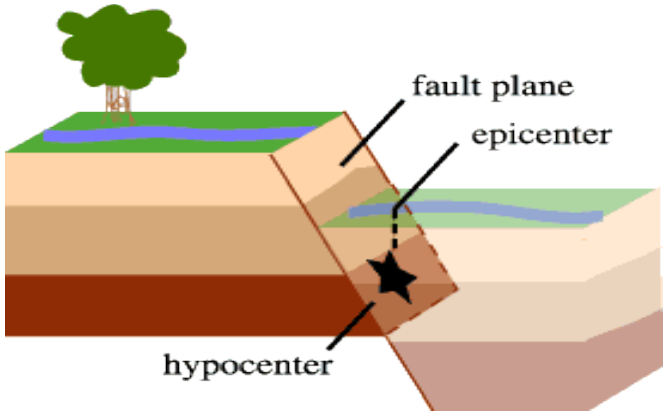


Figure 1. Earthquake visualization of Earth's Crust [26]

Deep underground is formed along plate boundaries due to tremors and occurrences of earthquakes. As the plates slide past and collide with each other tension and pressure build up and sometimes even glue together. Although the plate boundaries themselves may be glued, but there is a move and pull among the plates. Eventually, the demonstration of pulling is so much that the plates suddenly become brittle ruptures from each other, causing an earthquake!

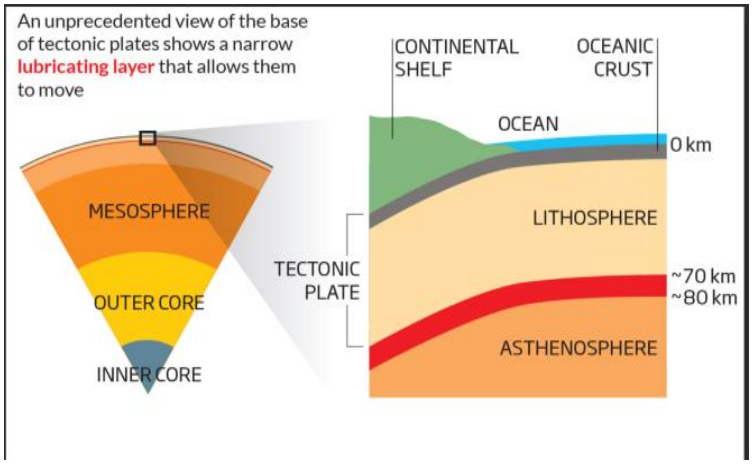


Figure 2. Movement of Earthquake and Seismic Waves [27]

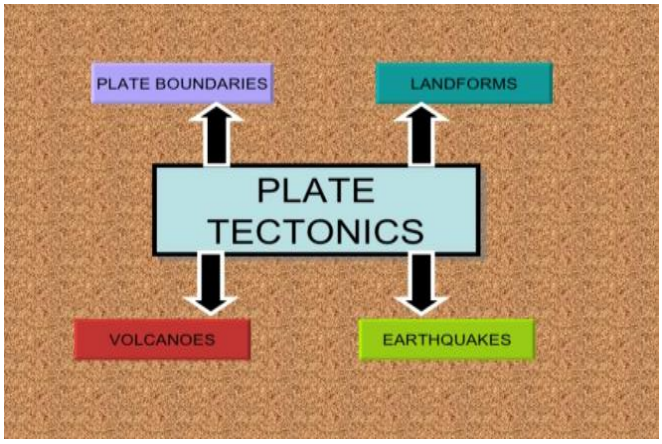


Figure 3. Different Tectonics Plates [28]

2. Literature Review

C.Henry, Das S. Huge, Superficial earthquakes the aftershock zones aftershock area expansions, scaling relations and fault dimensions to earthquakes, by Geophysical Journal International, 2001 [1]. This paper studied the hypocenters for relocation of earthquakes for dip-slip fault 64 and strike-slip fault 8s together. The data from Pegler & Das (1996) for strike-slip 27 earthquakes was also included in the studies. The location of the hypocenter was recorded, and their random strike was studied closely for both motion of earthquake in reverse direction and horizontal movement. It was found more frequently earthquakes appear at subduction zones and started the downwards dip edge of the fault.

Peresan A, Romashkova L, Kossobokov V, Panza G, Earthquake predictions and sustaining facts on a moderate-term midsized-range earthquake, by Earth and Planetary Sciences Journal, 2002 [2]. A Classification system considering SVR(support vector regression)and

HNN(hybrid neural network)was developed for obtaining the prediction of earthquake. To derive the relevant features, Minimum Redundancy (mRMR), Maximum Relevance were applied.

Jordan A, Thomas H, Jeffrey J McGuire, Li Zhao, Global Catalogue of Large Earthquakes giving information on predominance of unilateral rupture subduction zones or inland regions, by the Seismological Society of America Bulletin, 2002 [3]. How the rupture of an earthquake cultivated across a fault was studied and included in this work. It reflected on both the beginning and features of the fault and the rupture obtained due to the stress occurred continuously. Second moments of the earthquake through space-time distribution was used in this research. It quantified the propagation of an earthquake rupture.

Witold Dzwinel, David A.Yuen, Krzysztof Boryczko, Yehuda Ben-Zion, Shoichi Yoshioka and Takeo Ito, Landslide prediction based on support vector model, by IEEE, 2003 [4]. This paper was based on the earthquake patterns observed as a result of multi-resolutional scaling and clustering which were non-linear in nature. The implementation of these techniques were non-hierarchical and observed in synthetic seismic catalogues and then these techniques were investigated to get the novel result. The researcher explored the multitudinous earthquakes by covering a non-hierarchical clustering algorithm and non-linear multi-dimensional scaling with real time 3-D visualization and by inspecting the multivariate clusters.

W. S. Chen1, Q. Bai1, Q. Ren2,Z. Li1, Z. Y. Q. Xia1, Y. R. Liu1, A versatile and multifaceted method for huge earthquake prediction as there is a possibility of prediction of an enormous kind of earthquake, by Natural Hazards and Earth System Science, 2003 [5]. In China, in April 1997, the details studies were made in Xinjiang and prominent prediction was established based on the occurrences of earthquakes recorded in the different regions of that locality. On the basis of exhaustive analyses and usage of multiple innovative methods, results were collected. It viewed an ultra-low frequency range, effecting the infrasonic wave, crustal stress measurement, and abnormal behaviour of certain animals. These all were well recorded. Number of examples of prediction were also calculated which were thought successful in terms of prediction.

Aires, F. Bayesian statistics with application to remote sensing was applied to Neural network for uncertainty assessment, by Journal of Geophysical Research Atmospheres, 2004 [6]. An avenue to assess suspicious errors (i.e., the correlation between the structure of errors and error bars), were termed as "synaptic weights". This was based on the neural network variables including the principle of Bayesian technique and was presented in the first article. "Point estimation" weights in contrast to that of weights of the neural network (i.e., learning process can determine only one set of weights), were prescribed and assessed. This study monitored and noted the uncertainties of the neural network model on such estimates and the quality of data produced.

J. R. Holliday, K. Z. Nanjo, K. F. Tiampo, J. B. Rundle, and D. L. Turcotte**,** Forecasting of earthquake and its affirmations by Non-linear process in geophysics, 2005 [7]. To earthquake forecasting a modern approach including the concept on a pattern informatics (PI) method was studied in this paper, which quantified temporal variations in seismicity. With future occurrences of large earthquakes, derived from an association of small earthquakes, covering the areas in a seismogenic region known as hotspot. Hotspot was the location where earthquakes were forecasted to occur within the time span of 10 years in near future.

Allen RM, Olson EL, The deterministic nature of earthquake rupture, by Nature, 2005 [8]. In this paper, slip between plates were initiated on a small fault patch across the plane region between the faults and continued to rupture further was explained by the cascade model as long as the conditions for ruptures were strong and appropriate to occur. This domino-like theory had two elementary implications stated that small and large earthquakes begun or started occurring in the same manner and the second one was that of the rupture process was not deterministic, until the rupture stopped . Before the rupture was complete this model as it was difficult to determine the size of the earthquake estimated the magnitude of an earthquake.

J.B. Rundle, J.R. Holliday, D.L. Turcotte, K.Z. Nanjo, Forecasting of large earthquakes for optimal pattern informatics studied for Japan and its application by Pure and applied Geophysics ,2006 [9]. The faults underneath on which earthquake occurred faults interacted with each other, on a spatial and temporal scales considering large scale and were available underneath the tectonic plates. The two main features of regional seismicity were time and space and also other features appeared which were considered hypothetical in nature. The phase changes had been successfully compiled studied properly as related to concepts of

statistical physics to derive a variety of cellular automatic models . This paper was studied methodically to forecast the occurrence of earthquakes by studying the variations in seismicity.

Vincenzo Convertito, Antonio Emolo, Aldo Zollo, Seismic-Hazard assessment an integrated probabilistic–deterministic method, for a characteristic earthquake scenario by Bulletin of the Seismological Society of America, 2006 [10]. This paper was based on strong-motion parameter considered and chosen as location dependent variable for the evaluation of the probability of occurrences of earthquake exceeding than calculated. The statistical analysis of the synthetic waveform database occurring on the earthquake fault, was presented for a large number of possible responsible parameters and characterised their presence on rupture of the plates which occurred in past occurrences of earthquake. The threat was represented by the sliding plates of the earth due to collision or a fault having the similar symmetry ,shape and procedure found in the Umbria locality. This method was also applied to evaluate the hazard prone area of the same region where there was a possibility of earthquake occurrences.

Yih-Min Wu1 and Li Zhao, The early warning, the earthquake estimation of the magnitude was made and also studied the P-wave amplitude for first three seconds, by Solid Earth, 2006 [11]. Once in the first three seconds P wave arrived Pd represented peak amplitude of the displacement. Considering magnitude M, with the central distance R (hypothetically)the attenuation of Pd function was calculated based on earthquake in California at its Southern part. A relationship is formulated and expressed by the equation: log (Pd) = 3.463 + 0.729 M 1.374 log (R) ± 0.305. At stations close to the epicentre, a location of earthquake occurrence was decided and resolved by the P wave arrival timing. ‘‘Pd magnitude’’ of earthquakes was coined based on the above relationship.

Dreher Heinz., Immagadda Shastri L, Modelling and mining of earthquake data based on ontology principle, using data warehouse prediction analysis along Australian-Eurasian continental plates, by IEEE, 2007 [12]. In this paper, data entities were modelled based on data warehouse using ontology-base multidimensional representation. Authors have suggested that warehoused data modelling facilitates more effectively for earthquake prediction analysis .A multi-dimension in a warehousing environment was used to store these data structures to decrease the complexity of heterogeneous data.

James R. Holliday, Kristy F. Tiampo, Donald L. Turcotte, John B. Rundle, Chien-chih Chen, Forecasting earthquake principality or realm based on pattern informatics, by Seismological Research Letters, 2007 [13]. This paper showed that the observation was splitted into two general classes. The first one was based on empirical observations of precursory changes ,like change in seismic activity, ground motions, in temperature , atmospheric pressure and others precursors. The second approach included parameters like time, depth, magnitude etc. which was based on seismicity collection and inclusion on statistical patterns .Both the approaches were not very effective individually ranging days to months to years to provide reliable short-term forecasts but was fruitful when applied together on a basis of comparison between two approaches.

Alatas, B., Akin, E., Karci, A. Modenar, Multi-objective differential evolution algorithm for mining numeric association rules, by Applied Soft Computing, 2008 [14]. In this paper, for mining accurately differential evolution or DE an understandable and easy numeric AR or association rules algorithm was analysed and an advanced Pareto-based multi-objective search strategy was introduced which gave optimal results. The proposed study did not depend upon the support and the confidence thresholds considered smaller in effect rather it performed on independent database to get logical result which were hard to determine for each dataset used separately as single entity.

Zhong, Z., Yin, Y., Wang, Y., Quantitative association rules for data mining applying interval clustering, by Journal of Computational Information Systems, 2008 [15]. In this paper, to help in earthquakes prediction the modelling was performed on using regression techniques and the extraction was made on quantitative data on the carrier of seismic temporal data applying association rules. Thus, an attempt was made at explaining the conditions using a simple method for mining under which an occurrences of earthquake might occur referring the concept of the k–smallest , greatest values defined in association rule. Whereas the linear model, tree based model discovered , analysed and used the pattern generated by k values to find the occurrences of earthquakes.

Philip B. Stark ,Brad Luen, Testing earthquake Predictions, by Mathematical Statistical Institute, 2008 [16]. This work considered an earthquake of magnitude scheme of 5.5 and greater than that occurred anywhere in the world and forecasted. The prediction was made for the epicentre in a range covering a distance of 50 km. and concluded that a large earthquake would occur within 21 days. The rules applied in this work were to generate

different types of predictions and was included in Harvard centroid-moment-tensor (CMT) catalog in the year 2000-2004. Earthquake predictions had undergone a statistical test with a null hypothesis to study the chances and success rate of earthquake predictions.

Mehmet Karakose, Ilhan Aydin, Erhan Akin , Data mining based on time series method using fuzzy logic algorithm for the prediction named as prediction algorithm, by Semantic Scholar, 2009 [17]. The data mining techniques were created by a group of methods that revealed uncovered temporal patterns ,combined with fields of time series analysis which were treated as time series events characteristics utilized for the occurrences of earthquake or for earthquake prediction. In this paper a prediction algorithm based on time series data mining method was proposed applying fuzzy logic algorithm.The time series analysis was made by transforming time series data non-linear in nature and then to these data to predict the optimal values of the time series events parameters were considered fuzzy logic was applied for characterizing occurrences of earthquake.

Yan, X., Zhang, C., Zhang, S., For identifying association rules genetic algorithm-based strategy without specifying actual support in smallest quantity by Elsevier, Science Direct, 2009 [18]. This paper was designed and developed for identifying association rules without specifying actual nominal support based on a genetic algorithm-based strategy concept. A global search could be performed, and system automation was implemented, with the same genetic algorithm. This model did not require the user-specified threshold of support in smaller effect. This strategy was amplified to cover association rule in quantitative nature to identify the effect to be implemented on database for prediction.

Kwang Hoon Chi, Jin A.Lee, JongGyu Hang, Association rule was applied in quantitative manner, by of Earthquake data mining, by ACM Digital Library, 2009 [19]. The different relationship associated could be extracted from this study including components like location and frequency, intensity and timing, deepness, and intensity. The result of the study, enforced to apply the data mining technique to data of earthquake studies as the relationships of the components played a significant role. The association rule which was included quantitative mining on the data based on earthquake studies, also included date, effected area, deepness, intensity of earthquake and were put into small sectors to analyse effectively. This study applied a method to detect the association rule correctly and effectively by merging the occurrences of close by stations and applied repetitive process.

Hojjat Adeli, Ashif Panakkat, Recurrent Neural Network for approximate earthquake time and location prediction using multiple seismicity indicators, by Computer aided Civil Infrastructure Engineering, 2009 [20]. For predicting the occurrence of future moderate-to-large earthquakes two major parameters time and location in an approximate sense was considered named as computational approach using neural network modelling was presented. In this study a vector of eight seismicity was and considered as input parameters and used as indicators. For each time-period of following occurrences of major earthquake computation was done on seismicity indicators and their relationship was analyzed regarding the epicentral location having latitude and longitude and time of occurrence. This paper studied combining all the features by using a recurrent neural network.

Adeli H, Panakkat A., A chosen and selected neural network for earthquake magnitude prediction, by Neural networks, 2009 [21]. This paper was presented by PNN a probabilistic neural network using a predetermined future time -period in a seismic region for predicting the magnitude of the largest earthquake having eight mathematically computed parameters known as seismicity indicators. Three different statistical measures were evaluated by using the detection probability ,alarm ratio in fallacy, and the skill score or R score in tautology of regression analysis for accurate prediction and the effective model performance.

P S Koutsourelakis., Assessing structural vulnerability against earthquakes using multi-dimensional fragility surfaces, a Bayesian framework, probabilistic by Engineering Mechanics, Elsevier, 2010 [22]. A Bayesian framework was proposed which could produce estimates accurately for the derivation of brittle curves irrespective of the amount of data available to analyze. This paper discussed the flexibility particularly for the estimation when combined with MCMC or Markov Chain Monte Carlo techniques and could provide credible intervals of the curves effectively and accurately. A popular methodology principally was illustrated based on logistic regression that could lead to the derivation of brittle surfaces and curves. This model expressed the surety of exceeding a damage level with respect to several measures of the earthquake load, curves, and surfaces and could produce more accurate predictions of occurrences of earthquake.

J. Douglas Zechar. Earthquake forecasts and earthquake predictions were evaluated, by Statistical Seismicity Analysis, 2010 [23]. In various stages, to obtain quantitative answers to the questions using one or many methods which could be used: “Was earthquake predictions particularly good using these sets of questions”, “Was there a high probability to obtain the

same success?", "Were the records obtained and observed for earthquakes compatible with this earthquake forecast? "and "Which of these two forecasts was better?" Each evaluation record was analyzed in general terms having merits and demerits. These examples were chosen both specifically suited or using restricted or unusual readings. To refer an ideal example Poisson space-rate magnitude forecasts was referred. This study evaluated earthquake predictions and earthquake forecasts by reviewing various methods. It was understood by studying various available data examples and was accompanied by coding and compiling the source data.

Dr. A. Pethalakshmi a, R.Raja Rajeswa. Earthquakes Precursors and Prediction, by Journal of Computer Applications, 2011 [24]. This article not only explored the earlier efforts of earthquake prediction but also studied the role of soft computing in this realm and considered the appropriate precursors for earthquakes. Since two decades earthquake prediction research was being carried out though, still existed a challenge for the combination of earthquake prediction algorithms, Realtime systems and Geographical or spatial data mining.

S.Prakash, M.Vijayakumar and R.M.S.Parvathi, A novel method of mining association rule with multilevel concept hierarchy, by International Journal of Computer Applications, 2011 [25]. Many works were proposed for mining, though frequently the association rules were used to get the accurate result. The association rules both occurred unfrequently and regularly were more interesting ,easy and did not have duplicate or repeated rules concentrated mainly on mining data used for the research. The principle was based on the predefined rules formed with elements in the actual dataset processed and then compared with propositional logic applied on the same data set. The result obtained by association rules would not have duplication and they would be logically perfect to give output. General association rule would be applied if single level mining was performed on dataset. These association rules could be very useful in very high level decision making to get the logical result. In order to analyse the datasets for in-depth decision making, multiple level mining processes were employed with specific association rules for finding the dominant patterns.

3. Problem Statement

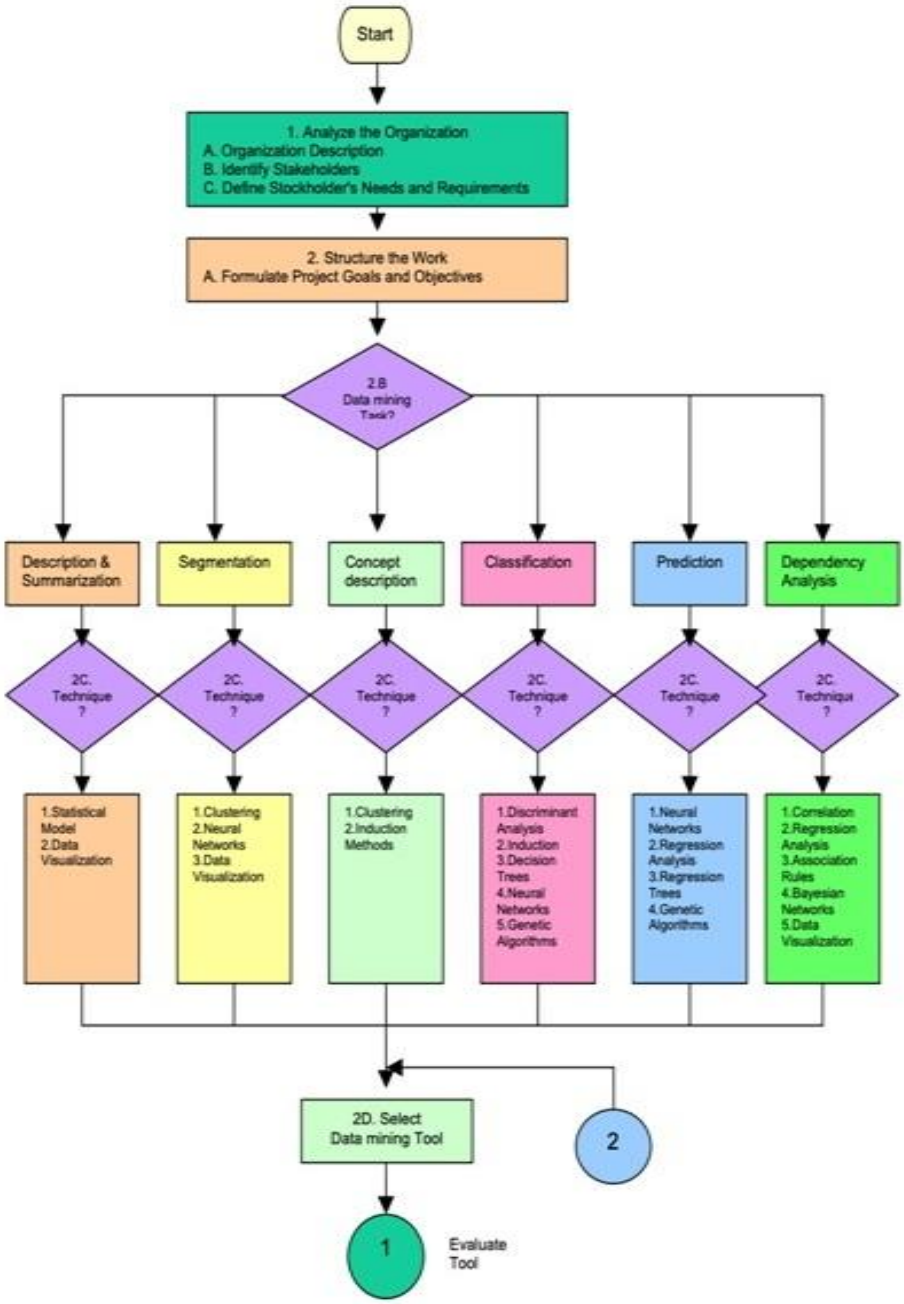
This research aims to design a scientific novel model that will be able to predict probable occurrences of earthquakes, utilizing the specified range of location having appropriate parameters and precursors.

4. Objectives:(Determinable of the Study)

1. To validate a proposed association on complete data on the earthquake record and apply data mining on the geophysical measurements.
2. To collect the statistical data and surveyor's data from different sources.
3. To employ the machine learning methodology for analyzing data so that it is applied on the statistical data on parameters like Depth, Time, Magnitude and Location etc.
4. To employ machine learning methodology on surveyor's data having precursors like variations in Temperature, Pressure, Animal Behavior, Falling of trees, water movements in river etc.
5. A novel model will be developed by corelating parameters and precursors to get appropriate prediction results.

5. Methodology

Data Mining is the most useful as well as modern concept an essential expository process intends to investigate information using a large database. An inter-disciplinary field of computer science uses data mining philosophy and its main aim is to discover patterns in large pools of data specified as big data. To find earlier unspecified facts , sensible patterns ,behaviour and relationships in huge data sets includes the application of defined and organized data mining tools. These tools can assimilate statistical models, machine learning techniques, and mathematical algorithms, as well as neural networks or decision trees. Data mining techniques involves several approaches association rules ,regression, clustering ,grouping rules, classification, summarization, pattern recognition .



.

Figure 4. Steps of Data Mining Concepts to be studied

Step 1: Seismograph Model

To study the occurrences of three parts of world a) the effected regions of India for most part of India where the occurrences of earthquake like b) the rift valley of Africa and c) Most parts and regions of Nepal, would like to meet the Geographical and Metrological Institutes and collect the data from all the Institutions, Specific Departments and NGOs of those regions.

To study the data gathered from Metrology Department which are generally recorded by leading Instruments like a) Global Positioning System or GPS the effective and efficient tool across the world for identifying the crust mobility. To detect the mobility of tectonic plates along the crust GPS is used, as GPS works in satellites. Using the GPS, the researchers and scientists can know the tectonic plates are in movable and rapid state, in order, to notify livelihood in those areas through recording different parameters like date, time, location, depth, magnitude, latitude, longitude etc. b) The Richter scale is also used to record the parameters of seismic activity having different parameters like depth, magnitude etc.

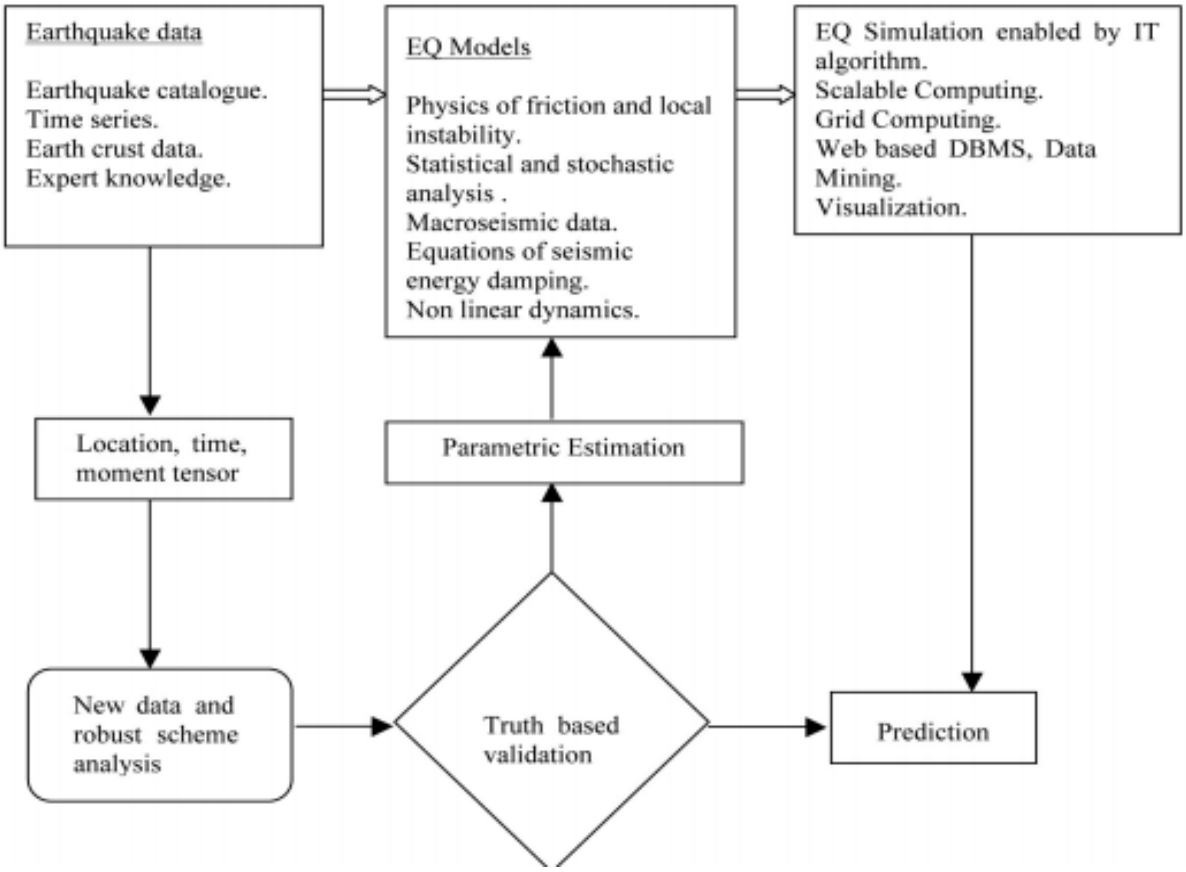


Figure 5. Seismograph Model

Step 2: Data Samples for Parameters

To study the data collected and the system and conclusion drawn from the above sources and then will like to collect data by survey form in different ways like observation, interaction with local people, questionnaire and finding details with the people who were actually affected in those areas during earthquake. Some people have experienced the earthquakes twice or more in lives. The survey forms were also filled telephonically and through emails.

For parameters we will choose date, time, longitude, latitude, magnitude, depth, region, location etc.

Step 3: Parameter Data Mining method

Once the big data are collected pre-processing is applied to extract the appropriate and suitable data by i) monitoring errors ii) standardize data format iii) validation on data accuracy iv) removing duplicate data v) analysing the data.

To apply association rule for pattern recognition, cluster analysis, clustering, neural network regression, random forest, time series etc. on the final Statistical data collected from Meterlogy Department using different parameters like date, time, depth magnitude etc.

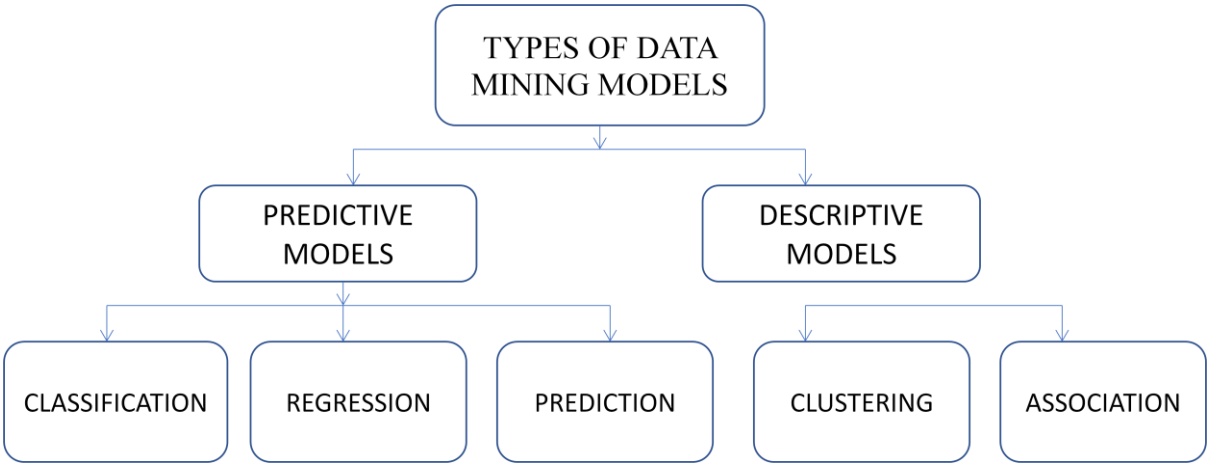


Figure 6. Basic Data Mining Method

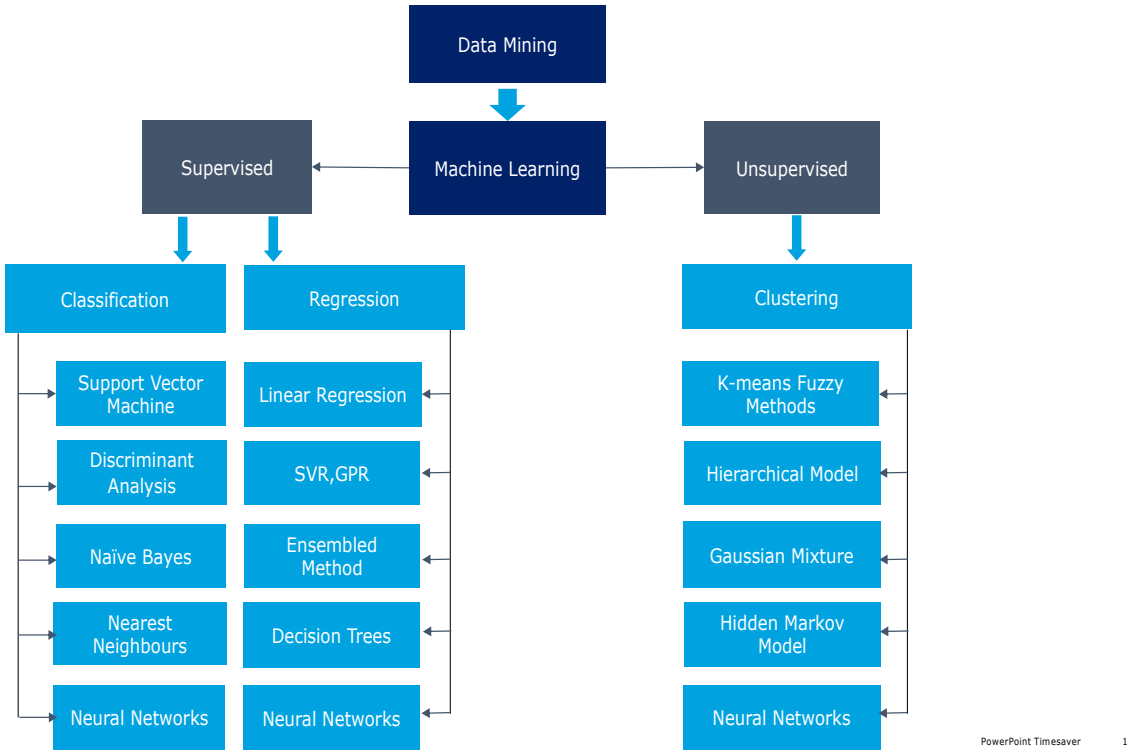


Figure 7. Different Machine Learning Methodology

Step 4: Precursor Survey Method

A survey form is designed having all the possible precursors responsible for occurrences of earthquake and taking all the surveyor's responses filling all the possible precursors like time, date, magnitude, behaviour of trees and plants, movements of animals, temperature, atmospheric pressure from the people actually experienced earthquakes by deploying for forecasting for the possibility and probability of occurrences of the earthquake. To apply association rule and clustering, decision tree, rule based methods, random forest, neural network on the precursors.

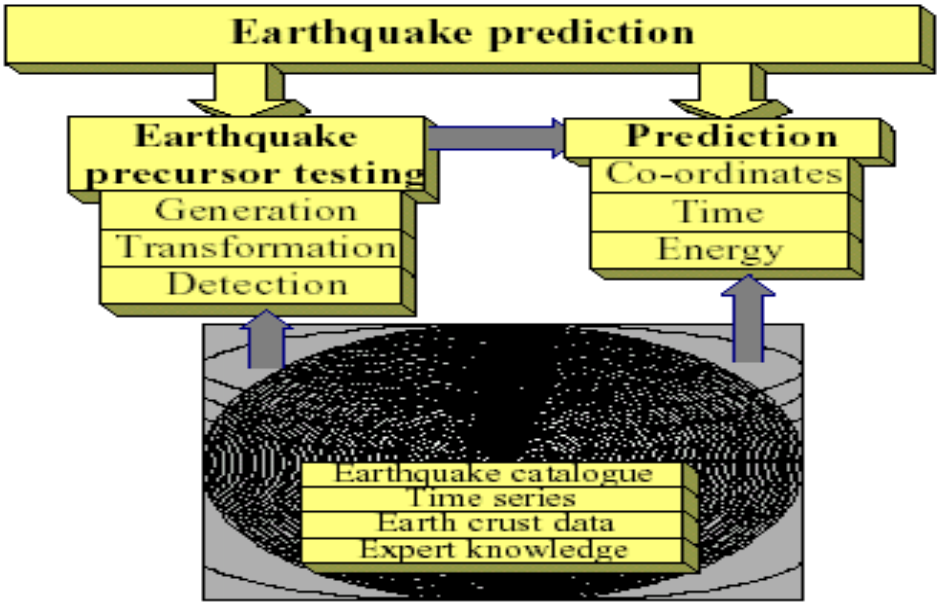


Figure 8. Precursors Details [29]

Step 5: Parameter Precursor Data Mapping Method

Data Mapping to be made on both kind of data analysis on parameters and precursors through a process of grouping similar data is known as clustering by analysing similar patterns through neural network.

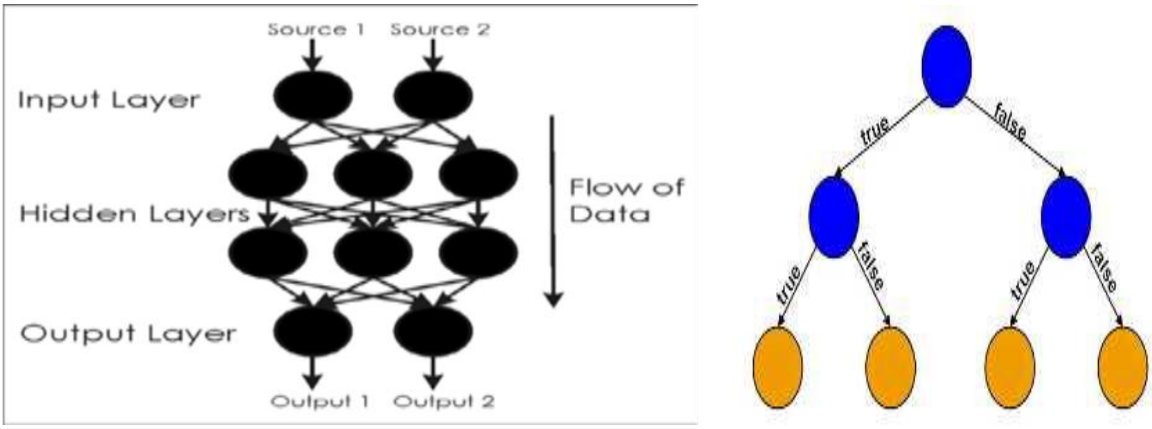


Figure 9. Data Mapping Method for both Parameters and Precursors [30]

7. Conclusions

We anticipate that in coming years demand of big data will be magnanimous and it will be strenuous to manage data by data handlers and researchers as well as to utilize both hardware and software support without modern technology application. Data mining extraction and implementation will apply its own techniques like neural networks, visualization, prediction etc. and will extract the big data in such a manner that only the productive, useful and applicable data are viewed from the big data pool to process easily, effectively, logically in an understandable and significant manner. The linear regression method and time series pattern on seismic waves which is applied both mathematically and statistically can increase the accuracy of predicting the magnitude and depth related to a specific location based on a probable zone. The overall R square was between 40% -50% for two different countries could suggest that the larger earthquakes occur at flat and hollow depths in the layers of crust of earth and that to in a particular zone. But tiny and minor earthquakes can and do take place at all depths and varied zones of different regions. These huge, broad and major earthquakes are associated with frictional sliding on faults, which can only occur at cooler temperature. Magnitude is stronger and positive impact at the time of early mornings and evenings in colder regions while it has comparatively lesser impact in late nights and hot regions. Depth has positive and significant impact in India while magnitude has positive impact in Nepal. Overall depth shows stronger impact than magnitude by comparing coefficients of both the model. Time series analysis is made on data of Nepal on Depth and Magnitude separately. The other analysis is made on data of India on Depth and Magnitude separately. Decision Tree analysis is carried on Surveyor's data. Basically, want to imbibe both parameters and precursors in such a manner so that one logical analysis is made to understand the occurrences of earthquake.

References:

[1] 2001, C. Henry, Das S. Huge. ,Superficial earthquakes the aftershock zones: aftershock area expansions, scaling relations and fault dimensions to earthquakes ,Geophysical Journal International, Volume 147, Issue 2, November 2001, Pages 272–293, https://doi.org/10.1046/j.1365-246X.2001.00522.x

[2] 2002, Kossobokov V, Romashkova L, Panza G, Peresan A. ,Earthquake predictions and sustaining facts on a moderate-term midsized-range earthquake, Annual Review of Earth and Planetary Sciences Vol. 30:1-33 (Volume publication date May 2002) https://doi.org/10.1146/annurev.earth.30.100301.083856

[3] 2002, Jeffrey J.McGuire, Li Zhao, Thomas H.Jordan A. ,Global Catalog of Large Earthquakes giving information on predominance of unilateral rupture subduction zones or inland regions, Bulletin of the Seismological Society of America 92(8):3309-3317 (December 2002|)DOI:10.1785/0120010293

[4] 2003,Witold Dzwinel, David A.Yuen, Krzyszt Boryczko. , Yehuda Ben-Zion, Shoichi Yoshioka and Takeo Ito, Landslide prediction based on support vector model, Print ISSN: 0974-2115 www.jchps.com Journal of Chemical and Pharmaceutical Sciences

[5] 2003, J. Z. Li1, Z. Q. Bai1, W. S. Chen1, Y. Q. Xia1, Y. R. Liu1, and Z. Q. Ren2, A versatile and multifaceted method for huge earthquake prediction as there is a possibility of prediction of an enormous kind of earthquake, Natural Hazards and Earth System Science, Copernicus Publications on behalf of the European Geosciences Union, 2003, 3 (6), pp.703-712. ffhal-002990

[6] 2004, Aires, F. Bayesian statistics with application to remote sensing was applied to Neural network for uncertainty assessment, Journal of geophysics and atmosphere ,21 May 2004, https://doi.org/10.1029/2003JD004173

[7] 2005, J. R. Holliday, K. Z. Nanjo, K. F. Tiampo, J. B. Rundle, and D. L. Turcotte**,** Earthquake forecasting and its verification, non-linear process in geophysics, volume 12,issue 6, Nonlin. Processes Geophys., 12, 965–977, 2005, https://doi.org/10.5194/npg-12-965-2005

[8] 2005, Olson EL, Allen RM., The deterministic nature of earthquake rupture, Nature ,438(7065):212-215, DOI: 10.1038/nature04214 PMID: 16281033

[9] 2006, K.Z. Nanjo, J.B. Rundle, J.R. Holliday, D.L. Turcotte, Pattern Informatics and its Application for Optimal Forecasting of Large Earthquakes in Japan, Pure and applied Geophysics ,Pure appl. geophys. 163 (2006) 2417–2432 0033–4553/06/122417–16 DOI 10.1007/s00024-006-0130-2

[10] 2006, Vincenzo Convertito, Antonio Emolo, Aldo Zollo, Seismic-Hazard Assessment for a Characteristic Earthquake Scenario: An Integrated Probabilistic–Deterministic Method, Bulletin of the Seismological Society of America (2006) 96 (2): 377–391. https://doi.org/10.1785/0120050024

[11] 2006, Yih-Min Wu1 and Li Zhao, The early warning, the earthquake estimation of the magnitude was made and also studied the P-wave amplitude for first three seconds, Solid earth, https://doi.org/10.1029/2006GL026871(24th August 2006)

[12] 2007, Immagadda Shastri L, Dreher Heinz. ,Ontology based data warehouse modelling and mining of earthquake data: prediction analysis along Eurasian–Australian continental plates, Conference: International Conference of IEEE in Industry Informatics at Vienna, Austria, DOI:10.1109/INDIN.2007.4384825(July 2007)

[13] 2007, James R. Holliday, Chien-chih Chen, Kristy F. Tiampo, John B. Rundle, Donald L. Turcotte, A RELM Earthquake Forecast Based on Pattern Informatics, Seismological Research Letters78 (1): 87–93. https://doi.org/10.1785/gssrl.78.1.87

[14] 2008, Alatas, B., Akin, E., Karci, A. Modenar, Multi-objective differential evolution algorithm for mining numeric association rules, Applied Soft Computing Volume 8 Issue 1, pp 646–656https://doi.org/10.1016/j.asoc.2007.05.003

[15] 2008, Yin, Y., Zhong, Z., Wang, Y.: Mining quantitative association rules by interval clustering, Journal of Computational Information Systems 4:609-616

[16] 2008,Brad Luen and Philip B. Stark Testing of prediction of earthquake was made by Mathematical Institute, 302-315 https://doi.org/10.1214/193940307000000509(2008)

[17] 2009 , Ilhan Aydin, Mehmet Karakose, Erhan Akin The prediction algorithm based on fuzzy logic using time series data mining method , Google Scholar, IEEE, Corpus id 8072525

[18] 2009, Yan, X., Zhang, C., Zhang, S.: Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support, Elsevier, Science Direct, www.elsevier.com/locate/eswa

[19] 2009, Jin A.Lee, Kwang Hoon Chi, Jong Gyu Hang, Mining quantitative association rule of earthquake data, ACM Data Library ICHIT '09: Proceedings of the 2009 International Conference on Hybrid Information Technology Pages 349–352https://doi.org/10.1145/1644993.1645059

[20] 2009, Ashif Panakkat, Hojjat Adeli, Recurrent Neural Network for Approximate Earthquake Time and Location Prediction Using Multiple Seismicity Indicators, Computer aided Civil Infrastructure Engineering, Volume: 24, Issue: 4, pp 280-292, DOI: 10.1111/J.1467-8667.2009.00595.X

[21] 2009,Adeli H, Panakkat A. A probabilistic neural network for earthquake magnitude prediction. By Neural networks. Sep;22(7):1018-24. doi: 10.1016/j.neunet.2009.05.003

[22] 2010, P S Koutsourelakis., Assessing structural vulnerability against earthquakes using multi-dimensional fragility surfaces: A Bayesian framework, Probabilistic By Engineering Mechanics, Elsevier, Volume 25, issue 1, pages 49-60, https://doi.org/10.1016/j.probengmech.2009.05.005

[23] 2010 ,J. Douglas Zechar. Earthquake forecasts and earthquake predictions were evaluated, By Statistical Seismicity Analysis doi:10.5078/corssa-77337879.

[24] 2011, Dr. A. Pethalakshmi a, R.RajaRajeswa. Earthquakes Precursors and Prediction, By Journal of Computer Applications, ISSN 0974-1925 ,volume 4,issue 2

[25] 2011,S.Prakash, M.Vijayakumar and R.M.S.Parvathi,A Novel Method of Mining Association Rule with Multilevel Concept Hierarchy By International Journal of Computer Applications ,ICACT(2):26-29.

[26] https://blog.geomodelr.com/

[27] https://www.newscientist.com/article/mg22530074-700-

[28] https://www.slideshare.net/maliadamit/

[29] https://www.ijsr.net/archive/v6i11/ART20177914.pdf

[30] http://www.webpages.ttu.edu/dleverin/. neural_network/neural_networks.html

MANAV RACHNA
Vidyanatarikshat

Swami Vivekananda Library and Resource Centre
Manav Rachna University

SIMILARITY VERIFICATION CERTIFICATE

Title: A Progression for the Prediction of Earthquake using Data Mining having an informative impact on Society

Researcher: Sumita Choudhary Registration Number: 2K17-PFE-014

Supervisor: Dr. PRINIHA Department: Computer Science & Technology

This is to certify that the above mentioned Short Synopsis/Thesis was scanned for similarity index through Turnitin Software on 1|10|21 with submission ID 1662330900 and the similarity index was found to be 0.6 %. The complete report running into 24 Pages, 5597 Word count with File Size 387.5 KBis being forwarded to the Research Supervisor/DRC/DRB for review.

Date: 1|10|21

Swami Vivekananda Library and Resource Centre, Faridabad ★ Manav Rachna University ★ 1/10/2021

The complete report of the above Short Synopsis/Thesis has been reviewed and the similarity index is below the accepted norms as per the University Grants Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions) Regulations 2018. The Short Synopsis/Thesis may be considered for further processing

Research Scholar

1/10/21
Research Supervisor

Sumita Choudhary Short Synopsis 1.10.2021

by Sumita Choudhary

Submission date: 01-Oct-2021 11:39AM (UTC+0530)
Submission ID: 1662330900
File name: A_Progression_Sumita_Choudhary_Short_Synopsis_1.10.2021.docx (847.92K)
Word count: 5597
Character count: 32832

A PROGRESSION FOR THE PREDICTION OF EARTHQUAKE USING DATA MINING HAVING AN INFORMATIVE IMPACT ON SOCIETY

SHORT SYNOPSIS

by

Ms. Sumita Choudhary
2K17-PFE-014

Under the supervision of

Dr. Prinima
Associate Professor, Department of Computer Science &Technology
Faculty of Engineering

Prof. Felix Musau
Dean of School of Computing, Riara University, Nairobi, Kenya

Abstract

Earthquake is a major debacle and entire world, and researchers are worried about this calamity. This calamity not only ascertains in injury, fatality of lives, and damages of buildings, property but also makes people helpless and homeless. This in turn, brings fright and forces people to concern about the causes of earthquake. The consequences of earthquake occurrences in near future are a dispute on lives of people globally. Earthquake forecasting is the demand of human race and a complex process to analyze dependency of various factors on the occurrences of the events. Precursors observed during earthquake can be treated as a parameter and through the studies and techniques of data mining; models can be prepared to show a correlation on the precursors and parameters. To analysis the problems and encourage the researcher these developed models are ideal to find essential solution to inherent in the

Sumita Choudhary Short Synopsis 1.10.2021

ORIGINALITY REPORT

Excluding matches
less than 14 words

6%	4%	5%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	pubs.geoscienceworld.org Internet Source	1%
2	Submitted to University of Nevada Reno Student Paper	1%
3	Yan, X.. "Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support", Expert Systems With Applications, 200903 Publication	1%
4	researchportal.hw.ac.uk Internet Source	1%
5	2dc40e33-085f-40e0-8172-9a1f898c1942.filesusr.com Internet Source	<1%
6	www.hindawi.com Internet Source	<1%
7	www.ijert.org Internet Source	<1%

8	W. Dzwinel. "Nonlinear multidimensional scaling and visualization of earthquake clusters over space, time and feature space", Nonlinear Processes in Geophysics, 01/28/2005 Publication	<1 %
9	Submitted to University of California, Los Angeles Student Paper	<1 %
10	citeseerx.ist.psu.edu Internet Source	<1 %
11	pubmed.ncbi.nlm.nih.gov Internet Source	<1 %
12	journals.plos.org Internet Source	<1 %

Exclude quotes	On	Exclude matches	< 14 words
Exclude bibliography	On		